VOLUME 61
NUMBER 5

WHOLE NO. 286
1947

Psychological Monographs

JOHN F. DASHIELL, *Editor*

Studies in Pilot Selection

I. THE PREDICTION OF SUCCESS IN LEARNING TO FLY LIGHT AIRCRAFT

G. GORHAM LANE
The Ohio State University

II. THE ABILITY TO PERCEIVE AND REACT DIFFERENTIALLY TO CONFIGURATIONAL CHANGES AS RELATED TO THE PILOTING OF LIGHT AIRCRAFT

RONALD R. GREENE
Ohio Wesleyan University

Published by
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Publications Office
1515 MASSACHUSETTS AVE., N.W., WASHINGTON 5, D.C.

TABLE OF CONTENTS

PREFACE

Data for this report are taken in part from research conducted at The Ohio State University under the auspices of the National Research Council Committee on Selection and Training of Aircraft Pilots, with funds provided by the Civil Aeronautics Administration.

Each of the authors is indebted to Dr. Floyd C. Dockeray of The Ohio State University for his valuable counsel and guidance in the designing and conducting of the experiments.

FOREWORD

UNTIL comparatively recently there has been little agreement as to what constitutes a successful airplane pilot. Likewise, although there have been many attempts to devise methods for the selection of individuals who if given proper training would be able to pilot airplanes with some degree of competence, the field has been only partially explored.

Psychological research in pilot selection began during the first world war. An excellent review of the type of work done during that period is given by Dockeray and Isaacs (3). Much of the early work was of doubtful direct value because of lack of opportunity for validation in actual flight training circumstances, and, in some cases, imperfect design, lack of adequate data and improper statistical treatment (18). It was, however, a definite contribution and pointed out, in many cases, the direction for later research.

After the cessation of hostilities interest waned, and by 1939 Jenkins (6) reports that no psychologist was working in the field of aviation psychology. Research in the field was not revived until the present conflict became imminent. However, in the interim between the two wars, laboratory work in the field of psychology was done, some of which, when the occasion arose, proved to be of value in the selection of pilots, even though originally it had not been undertaken with this in view. However, as had been the case with a large part of the early pilot selection research, the value of much of this was not immediately known because of a lack of opportunity to validate the findings in actual flight training circumstances.

It is recognized that the research done in recent years on analysis of piloting has resulted in a large amount of pertinent information. Likewise, research performed on the isolation of the personal characteristics which are involved in flying has yielded acceptable evidence as to their importance. Much of this information, however, has been related to military aviation as, for example, the work of Carlson (1) and Delucchi (2) where interest has centered in the selection of military pilots. With the exception of the research projects which have been done under the auspices of the National Research Council, Committee on Selection and Training of Aircraft Pilots (25), few investigations of a psychological nature have been concerned with the problems of the civilian aviator. In view of the widespread interest in private flying, and with the current liberal selection standards for candidates for the private pilot's license, it seems more than ever advisable to examine the selection procedures which are available. That this is a task for the psychologist has been expressed well by Kellum (11) who says that the real problem of the selection of aviators begins after the physical examination has been given. He points out that a large percentage of candidates who pass the physical examination eventually fail in flight training, and the reasons for failure are non-physical.

One of America's foremost investigators in the field of aviation selection during the recent war, Liljencrantz (14), defined pilot selection as follows:

> The process of selection may be conceived as consisting of the administration of a test or a single group or battery of tests, on the basis of which a dependable decision can be reached as to an applicant's aptitude for aviation.

The abilities required in successful piloting are best described as a complex of coordinations, skills, and abilities. Therefore, adequate pilot selection would best be accomplished through a method which combined various measurements of the components of this complex. Thus far, no combination of selection techniques has proved to be completely valid, nor has any test battery by itself or when combined with other selection techniques such as the informal interview or application blank, reached a point wherein its validity could not be improved. The question might be asked as to whether or not a test battery can be assembled to measure the various factors involved in learning to fly and if such a battery might be used to predict the success of candidates in learning to fly light airplanes. If any combination of selection tools can be made that serves the purpose of accurate prediction (of ultimate success or failure), this combination should prove to be more economical and valuable to use than any one of the components predictors by itself, or than the selection procedures already in use.

STUDIES IN PILOT SELECTION

I. Prediction of Success in Learning to Fly Light Aircraft

G. GORHAM LANE
The Ohio State University

STATEMENT OF THE PROBLEM

It was on the basis of the foregoing discussion that the present research was undertaken. A study was planned, first of all, which would make it possible to analyze some of the factors involved in the determination of success in learning to fly light aircraft. Various tests were available which had already been shown to measure factors involved in learning to fly, and a new test was added. This test was designed to measure factors not considered by any of the other tests. It was believed that these tests could be assembled into a battery not only which would measure some of the factors involved in learning to fly, but which would also be of some use in predicting success or failure in flight training. It was also believed that it would be possible to show how much each of the factors measured by the tests contributed to the prediction of success or failure in learning to fly. Finally, it was believed that the predictive value of the test battery might best be examined if several types of criteria were used. It seemed possible that a test battery might have high predictive value for a criterion of success in specific maneuvers, but be worthless in the prediction of over-all success or failure.

SUBJECTS USED IN THE INVESTIGATION

Thirty-seven male subjects between the ages of 17 and 29 years were used in this investigation. All were enrolled in the National Research Council flight training program at the Ohio State University, and received their flight training between the middle of October, 1945, and the first of March, 1946. Twenty-nine men were either enrolled in college, or had received bachelor's degrees and had research positions on the campus.

SELECTION TESTS USED IN THE PRESENT INVESTIGATION

All applicants for flight training under the experimental program were given a battery of tests. These tests were as follows:

1. The Self-Administering Test of Mental Ability
   (Gamma A. M. Otis quick-scoring)
2. The Ohio State Psychological Examination
   (Form 22)
3. Test of Aviation Information (Form P)
4. Biographical Inventory (Form 2C Civilian Key)
5. Test of Mechanical Comprehension (Form B, CAA)
6. Desire to Fly (Form XPA)
7. Mashburn Serial Reaction Test
8. Two-Hand Coordination Test
9. Judgment-Reaction Test

In the selection of tests for inclusion in this predictor battery, an attempt was made to include tests which were easy and economical to administer, closely related to aviation and which produced results capable of being interpreted in a straight-forward manner. In addition, the following considerations were utilized.

The role of general intelligence in flight success has been widely recognized (17), and the Test of Mental Ability of the C.A.A. fulfills the requirements of being a reliable instrument which can be used economically from the point of

view of both time and money. A test-retest measure reported for applicants for primary and secondary war service training under the C.A.A. is .79 (21). In administering the test, a twenty-minute time limit was used, and raw scores were utilized in statistical computation.

Since one study at least (10) has indicated that there is some relationship between grades in college and success in flight training, it was decided to include some such measure in the current battery. The N.R.C. Flight Training Course at the Ohio State University was open to non-students as well as to students; therefore, scholastic grades were not available for all individuals taking part in the study. However, the Ohio State Psychological Examination was available, and this test has been shown to have a correlation of .606 with first semester scholastic grades for male students (24). According to a verbal report made by Dr. Herbert Toops, author of this test, the reliability coefficients which have been calculated for this test are approximately .94. Grades are expressed in percentiles, based upon current norms at the Ohio State University. Percentiles were used in the statistical portion of this study.

The Test of Aviation Information (AI) was developed in 1941-43 at Wesleyan University and the University of Rochester, under the sponsorship of the Committee on Selection and Training of Aircraft pilots, NRC. Preliminary work done on this test indicates that "it can be included as one of the more promising predictors developed in the Committee's research program" (25). The test contains a total of 200 questions concerned with aviation. The raw score was used in all statistical computations. Reliability coefficients of .751 and .771 have been reported for this test (21).

The Biographical Inventory is otherwise known as the Inventory of Personal Data for Prospective Pilots (27). Regarding this inventory, Viteles has said (25):

> The Biographical Inventory represents one of the first, if not the first, successful attempt to predict pilot proficiency from biographical data.

Also, it should be remembered that biographical data have often been used in informal interviews. Psychiatric examination of candidates for flight training often includes questions dealing with the individual's past history. Johnson (8) has pointed out that biographical data, if properly used, are extremely valuable in the selection of individuals for aeronautical training.

A newly created Civilian key (The Kelly Positive Key) was used in scoring this inventory, and raw scores were used in the computation of statistical results. Reliability coefficients of .525 and .603 have been reported previously for this inventory (21).

Since, in any learning situation the factor of motivation is most important, it seemed desirable to include some measure of the student's interest in learning to fly. The Desire to Fly Inventory (12) was developed at the University of Rochester, and preliminary work on it has indicated that it has practical usefulness in a battery of predictors for success in flight training. It contains a total of 235 questions pertinent to interest in flying which are to be answered by the applicant. Key A B was used in scoring, and raw scores were utilized in the current investigation.

Although no reliability coefficients are available for the Desire to Fly Inventory, the authors indicate that it is reliable in their report to the National Research Council (12). In this report they state:

An analysis of the distributions of items answered "No" by various percentages of the populations of both samples A and B indicated that the items are fairly stable in the sense that each was answered "No" by approximately the same proportion of cases in each sample.

The Mashburn Serial Reaction Apparatus is one of the older psychomotor tests still being used in pilot selection. Actually the test measures the individual's ability to make rapid eye-hand-foot reactions. A complete description of this apparatus is available in McFarland's report No. 34 for the Airman Development Division of the C.A.A. (15). The total time required for making a series of forty eye-hand-foot responses was obtained and used in later statistical work. In the Boston-Midwest Study (20) reliability coefficients of .53, .74, and .74 were found for this test when three different samples were used.

The Two-Hand Coordination Test was developed from two tests formerly used in industrial selection, namely, the Wisconsin Miniature Engine-Lathe Test, and the Farmer-Chamber's Coordination Test. A complete description of the current version of the test is available (16). On this test the subject is given six trials and is scored on the percentage of time he maintains contact between two moving discs. The preliminary studies on this test indicate that the mean of six trials was found to correlate .78, .87, .80 with trials 4, 5, and 6. Therefore, the mean score for each applicant was used in this study. Reliability data are available from the Boston Midwest Study for three samples and were found to be .75, .50, and .80 (20).

The Test of Mechanical Comprehension (MC) contains seventy-six questions with companion diagrams concerning mechanical problems. Reliability coefficients of .697, and .743 have been reported (21).

In view of the fact that, in the past, tests involving reaction time have shown relationships greater than chance to flying success, it seems that any battery of tests constructed for the purpose of predicting success in flight training should be heavily weighted with tests involving reaction time. Therefore, in the present investigation, a new test was included, which, although involving a measurement of reaction time, was improved from the points of view of economy of construction, and of simplicity of administration and scoring. This has been designated as the Judgment-Reaction Test.

The Judgment-Reaction Test was based upon an apparatus originally designed by Ranschburg and called by him a "Mnemometer." For purposes of the present experiment the basic apparatus was modified and improved. As used in this research, the test consists of a black wooden box approximately 15 inches square in which the revised Ranschburg rotation mechanism was placed. The front surface of the box contains an aperture through which the stimulus is visible and two hand switches which are operated by the subject.

Stimuli consist of sixty shades of Hering grays so arranged on a cardboard disc that as the correct switch is pressed by the subject, the disc moves one-sixtieth of its circumference and reveals a new stimulus through the aperture. The subject is required to make a judgment as to whether the shade of gray which has become visible is darker or lighter than the one which was visible previously: If it is darker, he must press the switch on his right; if it is lighter, he must press the switch on his left. If he makes an incor-

rect response, the stimulus does not change. Only the preferred hand is used.

The subject is required to make a series of sixty judgment reactions and is scored on the quickness with which he completes the series. He is given a series of six trials of sixty judgments each with a rest pause of one minute between trials. In preliminary experiments done with this test, a correlation of .90 was found between trials 4 and 5. This was taken as the point at which the effects of initial practice were at a minimum and also as an indication of test-retest reliability. The time required to complete sixty judgment reactions in the fifth trial was used as the score for the subject in this investigation.

CRITERIA USED IN THE PRESENT INVESTIGATION

The criteria of success in flight training available in this study were divided into two main types: gross criteria, or measures of over-all success or failure in flight training; and specific criteria, such as observations of good or poor performance on specific aspects of flight performance. These are not mutually exclusive for it is entirely possible that a student might perform above average on most maneuvers, yet not be able to combine their performance into a smooth pattern and, thus, be rated as a "poor" flier. Still again, a student might be able to perform maneuvers and specific aspects of flight well, yet be labeled an "unsafe" pilot. On the other hand, a student might successfully pass his private flight examination by giving an overall good performance, although inspection of his individual grades might reveal that he was better in some aspects of flight performance than in others.

There is also a large variation in the student's performance from day to day, so it seemed advisable that any data collected on his performance should be sampled over a period of time in order to obtain as accurate a picture of his flight performance as possible.

Furthermore, there is the ever present difficulty of unreliability of criteria obtained from ratings. One way by which it is possible to minimize the error due to individual ratings is to combine ratings or to secure ratings from different observers whenever possible.

Another consideration in the selection of criteria, especially in a study such as this one involving flight performance, is the practicality of collecting observations. This factor necessarily limited the types of observations which could be collected.

Keeping in mind these principles, criterion data were obtained from four sources:

1. The C.A.A. Flight Inspector
2. The student's instructor
3. The check flight pilot
4. The student's log book

Two measures were available from the C.A.A. Flight Inspector. These were the overall grades on the private pilot test and the demerit score given on the student's landing performance.

The C.A.A. grade on the private pilot test represents the inspector's opinion of the student's skill in piloting. It determines, in part, whether or not the student is issued the private pilot's license. Normally, this test, which comprises a series of maneuvers prescribed by the Civil Aeronautics Administration, does not take place until the student's instructor has recommended him as being ready for the private pilot's license. However, in the experimental program, all students received this test after they had completed thirty-five hours of flight training

Previous research has shown that the inspector's grade on this test has a substantial relationship to some of the predictors which were used in the current research (10).

It was mentioned previously that in this study the attempt was made to sample criteria of two types, gross and specific. Since landing is one of the most complex and difficult of all maneuvers encountered in learning to fly, it was selected to represent the specific criteria. The following quotation from an Army Air Forces article expresses well the reason for stressing the student's ability to land a plane (23).

> Experts are of the definite opinion that landing in the proper place in the proper attitude without dropping the plane in or bouncing it involves important aspects of flying skill, namely, the ability to judge space and plan a course through it, to control the attitude and airspeed of the plane, and to feel when it is about to stall.

One requirement of the private pilot test is that the student attempt three spot landings; that is, attempt three times to set the plane on the ground within 300 feet of a designated spot on the runway. Failure to do this in two out of three attempts means failure in the entire examination.

According to C.A.A. regulations, scores on individual maneuvers such as this are given on a demerit basis. Demerit scores range from one through five, having values as follows:

1. Excellent (90-100)
2. Above average (85-90)
3. Average (80-85)
4. Below Average (70-80)
5. Unsatisfactory (0-70)*

Although the student has to attempt three landings, only one demerit score, representing the average score on his performance is given for landing. This score plus the overall grade on the private pilot test were selected to represent the C.A.A. inspector's opinion of the student's flying ability.

Five measures were available from the student's flight instructor and were included in the initial phases of this study. The first four of these measures represented the instructor's appraisal of the student's skill in landing. The instructor kept a daily check sheet on the student's performance during the flight lesson, and each maneuver performed was graded on a percentage basis, with 70 being regarded as passing. In order to obtain landing data on each student from the flight instructor, the following procedure was devised.

During each flight lesson students might make a variable number of landings. In some flight lessons, he might make none at all without assistance. In order to obtain an adequate sample of the instructor's grades on unassisted landings for each student, an average grade was computed for landings made by the student during the two flights immediately preceding each of the four check flights. This represented the most satisfactory method of obtaining instructor data on landings and these data are comparable on a time basis to that obtained from the check pilots.

Further data were available from the instructor in the form of ratings on a "Scale for Rating Pilot Competency." This scale was developed at Purdue University (9) and a factor analysis of the scale has indicated that the 14 items in the scale measure three distinct factors, tentatively called "skill," "judgment," and "emotional control." Preliminary work on the scale has indicated that its

* C.A.A. form ACA 342

use differentiates between the "best" and "poorest" students of a large group of instructors from several different areas.

There is a possibility of scoring each item on 40 points. Since the original experimentation done on this scale indicated that the most reliable results were obtained by adding the scores on the three factors, and combining them into three separate scores, this method was followed in the present study. The total number of points on each factor received by a student was regarded as his score.

The Ohio State Flight Inventory provided a major portion of the criteria used in this study and comprised all of the data received from the check pilots. The OSFI, as it is commonly called, was the result of research initiated at the Ohio State University in 1939, in an attempt to devise a standardized rating technique for use in making observations of student pilot performance (4). The most recent version of the OSFI was used (19).

This inventory was administered four times to each student during the course of flight instruction. It was used during the check flights which occurred at the end of seven, fifteen, twenty-five, and thirty-five hours of flight instruction. Two check pilots were used to administer this inventory, and these pilots flew with each student on alternate check flights. The inventory was filled out during the flight so that omissions due to forgetfulness or oversight might be kept at a minimum. The check pilots flew with the student only during the check flight, and thus, this opinion of the student's flight performance represents an independent estimate.

Two types of data were available from the OSFI. These are referred to as demerit scores and maneuver grades. Demerit scores refer to scores based upon individual aspects of a maneuver, and grades represent an over-all estimate of maneuver performance.

Each check flight required three ratings of the student's performance in making a "Final Approach and Landing." Data on this maneuver were divided into three main categories, "Control of Plane," "Precision," and "Safety." Each of these was broken down into subdivisions which could be scored separately. In this way it was possible to secure a detailed account of the student's performance on that particular landing. Scores on this check sheet were obtained in the form of demerit weights, ranging from one through three. If a particular performance was satisfactory, no score was given.

For the present purposes, the total demerit score for each landing was computed. Then an average was taken of the demerit scores on as many landings as the student made. In no case was more than one landing omitted, and this omission was generally due to incomplete scoring because the check pilot had to take over the controls. Average demerit scores were available on four check flights for each student used in the study.

Each landing that the student made during the check flight was also assigned an over-all grade by the check pilot. This grade reflected his opinion of the student's over-all performance in landing the plane. In all check flights the check pilot was asked to grade the student on a percentage basis, using 70 as the fixed grade for a passing score. At each stage in his training, the student was compared with the performance expected of a certified private pilot; in other words, he was graded on a fixed scale. The average over-all grade for each landing made was com-

puted for each of the four check flights.

In addition to making three regular landings during each check flight, the student was asked to execute one "Landing with a Slip." Performance on this maneuver was first scored on a demerit basis and the total number of demerits was computed for each "Landing with a Slip," making available a total of four sets of demerit scores. Over-all grades for this maneuver were also available and were included.

On the fourth check flight, at the conclusion of thirty-five hours of flight instruction, the student was required to proceed to a strange field and make a landing. This necessitated a somewhat different approach to the field since instructions were to make a "power on landing," whereas the other landings were made "power-off." Demerit scores for this maneuver were available for the fourth check flight only and were included in the criteria.

An independent gross criterion was obtained by keeping a record of the length of time required before the student was allowed to solo. Although popularly, such time measures as this have been thought to be of significance, previous studies have shown that this particular measure may not be too significant (25). "Time to Solo" records were available for all students in terms of hours and minutes before soloing, and were included as criteria.

Although the reliability of the criteria used in this investigation is still being studied in other investigations, some data are available which indicate that these criteria are suitable for use in connection with pilot selection.

In a study reported by the National Research Council (19), correlations between inspector's flight grades and the OSFI summation score were found to range from .83 to .95. This indicates a marked consistency between the ratings secured from these two sources. In a more recent report, Wapner and Bakan (26) have attempted to demonstrate the relationship of inspector's grades on the OSFI to photographic records of the student's performance. Their conclusions were that although, in general, the accuracy of the inspector's rating was high, there was some variation in the degree of accuracy for certain maneuvers. However, they feel that their study supplies empirical evidence to justify placing confidence in the inspector's criterion measures. Actually, this shows that such ratings can be accurate. And thus it seems justifiable to use raters who have been well trained in using the OSFI, in research on pilot selection.

In an analysis done by Johnson and Boots (7), of ratings in the preliminary phase of the CAA Training Program, it was found that correlations between inspector's final ratings and instructor's mean ratings on given maneuvers were low, even for the last two hours of flight. Furthermore, the intercorrelations between instructor's mean ratings for given maneuvers ranged from .29 to .93, with the higher correlations tending to be between maneuvers most frequently rated.

RESULTS AND DISCUSSION OF RESULTS

The procedure usually adopted when it is desired to evaluate the relative contribution of single tests to the predictive value of a battery of tests is to calculate regression coefficients for the tests involved. The method which is regarded as most satisfactory when working with many variables is that originally devised by Doolittle. A complete description of this method is available in Peters and

Van Voorhis (22) pp. 226-234. This technique was utilized in this study.

Each test was selected for inclusion in the battery because it was believed to measure some factor or factors involved in the ability to fly light aircraft. Therefore, some measure of the relationships between the tests was necessary, and intercorrelations were computed. Table I shows the zero order correlations among all of the tests included in the battery of predictors. In general, these correlation coefficients were not high, indicating that each test was fairly independent of the others. The major exception to this was the correlation of .744 between the O.S.P.E. and the test of Mental Ability. This might have been expected, however, from the nature of the tests. Next to this was the correlation of .625 between the two coordination tests, and this, too, might have been expected from the nature of the two tests involved.

Table II shows the intercorrelations among the criteria. As in the case of the predictors, the intercorrelations were, in general, low. The highest correlation was .725, between the same instructor's "scores" and "grades" on landings with a slip on the first check flight. Sixteen of the correlations were above .50. For the most part the higher correlations were found between grades and scores given by the same pilot on the same maneuver, thus giving an indication of the reliability of these ratings. Table III, below, shows the correlations which obtained between demerit scores and grades on landings when both measures were obtained from the same instructor.

All variables were coded in such a way that a positive correlation indicates a positive relationship between proficiency in both variables, with the exception of the C.A.A. landing grade. Negative correlation in this case indicates a positive relationship in proficiency.

The correlation between the two measures from the C.A.A. Inspector's flight test was −.512. A relationship such as this might be expected as the landing grade is one of the most important determinants of the over-all grade. Evidence that a positive relationship does obtain between these two grades is furnished in a study done on the analysis of Inspector's ratings by Festinger (5) in which it was found that among the most uniformly high correlations between mean maneuver grades and over-all grades on the private pilot flight test were those obtained from precision landings. These

TABLE I
Intercorrelations of Predictors

	Judgment reaction	Mental ability	OSPE	A.I.	B.I.	M.C.	D.F.	Mash.	Two-hand
Judgment reaction		.315	.353	.289	−.005	−.047	.203	.228	.270
Mental ability			.744	.420	−.017	.282	−.190	.376	.354
OSPE				.439	.062	.285	−.079	.383	.398
A.I.					.395	.419	−.208	.168	.276
B.I.						.204	.180	.066	−.013
M.C.							.087	.411	.526
D.F.								.145	.067
Mash.									.625
Two-hand									

TABLE II
Intercorrelations Among Criteria

	C.A.A. Over-all	C.A.A. Landing	Purdue Scale Skill	Purdue Scale Judg-ment	Purdue Scale Emot. cont.	OSPI Landing Scores 1	OSPI Landing Scores 2	OSPI Landing Scores 3	OSPI Landing Scores 4	Strange field	Instructor Grade 1	Instructor Grade 2	Instructor Grade 3	Instructor Grade 4
C.A.A. Overall		-.512	.686	.203	.481	.132	.319	.109	.498	.065	.379	.195	.190	.221
C.A.A. Landing			-.326	.000	-.110	-.330	-.200	-.171	-.065	-.151	-.271	-.090	-.093	-.108
Purdue Scale Skill				.251	.640	.255	.573	.376	.557	.189	.520	.413	.349	.367
Purdue Scale Judgment					.144	-.163	-.062	-.245	.065	-.157	.071	-.247	.067	.135
Purdue Scale Emotional Control						.162	.538	.450	.442	.221	.263	.071	.247	.306
OSPI Landing Scores 1							.179	.303	-.213	.102	.082	.173	.073	.071
OSPI Landing Scores 2								.355	.400	.296	.179	.405	.461	.122
OSPI Landing Scores 3									.183	.105	.519	.274	.142	.340
OSPI Landing Scores 4										.264	.399	.449	.378	.140
Strange Field											.182	.284	.247	-.142
Instructor grades 1												.351	.244	.194
Instructor grades 2													.539	.209
Instructor grades 3														.255
Instructor grades 4														

	Time to solo	OSPI landing grade 1	OSPI landing grade 2	OSPI landing grade 3	OSPI landing grade 4	Slip demerit scores 1	Slip demerit scores 2	Slip demerit scores 3	Slip demerit scores 4	Slip grades 1	Slip grades 2	Slip grades 3	Slip grades 4
C.A.A. Overall	.570	.045	.223	.366	.364	.059	.219	.230	.382	.190	.184	.274	.480
C.A.A. Landing	-.305	-.109	-.114	-.246	-.157	-.035	-.016	-.074	-.084	-.108	.068	-.064	-.189
Purdue Scale Skill	.670	.289	.478	.420	.293	.119	.614	.301	.344	.301	.434	.386	.431
Purdue Scale Judgment	.094	-.050	.066	.033	-.014	.036	.177	.013	-.078	.182	.237	.112	-.097
Purdue Scale Emotional Control	.201	.222	.383	.473	-.008	-.095	.549	.128	.368	-.017	.532	.125	.204
OSPI Landing Scores 1	.225	.409	.436	.411	-.226	.303	.200	.355	.186	.206	.273	.224	.123
OSPI Landing Scores 2	.427	.287	.580	.289	.203	-.043	.510	.263	.294	.073	.174	.238	.356
OSPI Landing Scores 3	.390	.297	.406	.634	-.238	.077	.353	.406	.198	-.030	.310	.249	.004
OSPI Landing Scores 4	.504	.173	.191	.074	.662	.078	.473	.042	.489	.187	.198	.182	.554
Strange Field	.223	.200	.232	.107	.342	-.049	.202	-.140	.460	-.104	-.075	-.190	.438
Instructor grades 1	.609	.062	.347	.442	.195	.148	.293	.178	.258	.176	.238	.211	.252
Instructor grades 2	.593	.330	.368	-.245	.359	-.050	.482	.277	.280	.015	-.076	.090	.386
Instructor grades 3	.384	.357	.271	.067	.234	.032	.514	.202	.213	.178	.139	.118	.383
Instructor grades 4	.111	.219	.136	.120	-.100	-.199	.367	.094	-.003	-.003	.233	.008	-.341
Time to solo		.360	.448	.367	.341	.262	.486	.479	.283	.370	.135	.303	.486
OSPI Landing grades 1			.375	.326	.024	.221	.548	.318	.312	.409	.291	.161	.178
OSPI Landing grades 2				.509	-.123	.176	.454	.496	.190	.265	.536	.479	.121
OSPI Landing grades 3					-.317	.313	.206	.535	.221	.267	.416	.495	.043
OSPI Landing grades 4						-.051	.215	.304	.480	.120	-.184	-.141	.725
Slip Demerit Scores 1							-.187	.304	.141	.725	-.012	.348	.272
Slip Demerit Scores 2								.211	.286	.158	.538	.128	.258
Slip Demerit Scores 3									-.127	.276	.261	.721	-.108
Slip Demerit Scores 4										.103	.127	.083	.678
Slip Grades 1											.170	.268	.235
Slip Grades 2												.274	-.038
Slip Grades 3													-.018
Slip Grades 4													

TABLE III
Correlations Between Demerit Scores and Grades on Landings (Same Instructor)

	Landing grades
First check	.409
Second check	.580
Third check	.634
Fourth check	.662

correlations are given below:

	Inspector C	*Inspector D*
Overall: Landing	.68 N = 27	.65 N = 28
Overall: Landing	.57 N = 28	.73 N = 28

Although two of the factors contained in the Purdue Rating Scale showed low positive relationship with one another, the correlation between ratings on "Emotional Control" and "Skill' was exceptionally high. It is quite possible that an emotionally stable individual would not be tense in handling the controls of a plane, and it is well known that much of the flight instructor's time is spent in attempting to get students to relax. Therefore, it is probably to be expected that ratings in skill and emotional control would be positively related.

The correlations between the demerit scores given on "Landings" in the four check flights were not high, indicating that probably the factors involved were fairly independent of one another. It is probable that landing performance actually changes as the student has more flight instruction.

The correlations between demerit scores on "Strange Field Landings" and other ratings were low with the exception of those between demerit scores and grades on "Landings with a Slip" in the fourth check flight. Actually this affords an indication of the reliability of the ratings as "Strange Field Landings" were administered only in the fourth check flight. Thus the same individual was responsible for these two ratings. The high correlation might also indicate that the same type of abilities are involved in making landings with a slip and strange field landings.

The correlations between instructor's grades on landings were all positive but not high. The highest was a correlation of .539 between the average grades on landings preceding the second and third check flights. The highest correlation between instructor's grades and any of the other criteria was one of .593 obtained between instructor's grades on "Landings" preceding the second check flight and "Time to Solo." It would appear that favorable ratings on landing performance are related to the time when the instructor gives the student permission to solo.

"Time to Solo" showed a fairly high correlation with the C.A.A. inspector's over-all grade, and a positive relationship with the C.A.A. landing grade. "Time to Solo" is most highly correlated with the instructor's rating of skill on the Purdue scale. This again suggests that the instructor actually does allow a student whom he feels is skillful, to solo early in the flight course. In general, this time measure showed fairly high correlations with the other criteria.

The intercorrelations between check pilot's grades on landings in the four check flights were all positive with the exception of ratings on the fourth check flight. In fact, the correlations between grades on landings in the fourth check flight show a progressively inverse relationship to grades on the first, second, and third check flight.

This is quite different from what might have been expected, since, it seems more logical that grades on landings should show more relationship to one

Table IV
Correlations Between Individual Predictors and Criteria

Predictors	C.A.A. overall	C.A.A. landing	Purdue Scale skill	Purdue Scale judgment	Purdue Scale emotional control	OSPI landing scores 1	OSPI landing scores 2	OSPI landing scores 3	OSPI landing scores 4	strange field	Instructor grade 1	Instructor grade 2	Instructor grade 3	Instructor grade 4	time to solo	OSPI landing grade 1	OSPI landing grade 2	OSPI landing grade 3	OSPI landing grade 4	Slip demerit scores 1	Slip demerit scores 2	Slip demerit scores 3	Slip demerit scores 4	Slip grades 1	Slip grades 2	Slip grades 3	Slip grades 4
J.R.	.156	.125	.216	.250	.289	-.239	-.180	.139	.298	-.151	.212	.050	-.146	.229	.076	-.011	.099	.099	.060	-.018	.269	.029	.125	.143	.193	.144	-.017
Otis	-.141	.148	-.115	.232	-.213	-.058	-.228	-.148	.001	-.260	-.239	.196	.144	.114	-.018	.174	-.117	-.365	.080	.067	.187	.126	-.196	.182	-.071	.033	.048
OSPE	-.016	.111	-.041	.306	-.089	-.285	-.110	-.100	.193	-.236	-.133	.125	.105	.095	-.017	.050	.018	-.212	.080	-.097	.176	.136	-.138	.059	.112	.071	-.040
A.I.	.351	-.356	.303	.027	.167	.130	.058	.126	.213	.145	.146	.167	.160	.100	.271	.319	.128	.152	.287	-.050	.271	.257	.143	.230	.138	.167	.346
B.I.	.302	-.309	.217	-.160	.191	.351	.241	.379	-.095	-.085	.197	.097	.390	.107	.144	.114	.411	.474	-.190	.096	.084	.405	.048	.112	.312	.371	.068
M.C.	.315	-.052	.133	.112	.222	.140	.136	-.185	.170	-.034	-.296	.023	.132	-.053	.031	.084	.075	-.038	.240	-.230	.147	.052	.118	-.041	.140	-.039	.232
D.F.	.414	-.122	.275	.037	.417	.043	.247	.273	.243	-.225	.345	.187	.072	.083	.209	-.174	.313	.206	.045	-.117	.132	.084	.171	-.174	.180	.209	.082
Mech.	.053	.247	.103	.204	.178	-.083	-.083	.028	.069	-.216	-.178	.102	.027	.200	-.120	-.160	-.034	-.199	-.101	-.282	.151	.087	-.109	-.376	.179	.032	-.152
Two-H.	-.016	.276	-.114	.226	.142	-.278	-.183	-.222	.152	-.100	-.273	-.182	.135	.167	-.267	-.087	-.188	-.290	.127	-.311	.121	-.291	.004	-.185	.130	-.237	-.060

TABLE V

Correlations of Predictors with Selected Criteria

Predictors	C.A.A.		Emot. Control	OSFI Land Scores		Instruct. Land Grades		OSFI Land Grades		Slip Demerits	Slip Grades		
	Overall	Land		1	3	1	3	2	3	3	1	3	4
Lane	.155	.125	.289	−.239	.139	.212	−.146	.099	.099	.029	.143	.144	−.017
Otis	−.141	.148	−.213	−.058	−.148	−.239	.144	−.117	−.365	.126	.182	.033	.048
OSPE	−.016	.111	−.089	−.285	−.100	−.133	.105	.108	−.212	.136	.059	.071	−.040
A.I.	.351	−.356	.167	.130	.126	.146	.160	.128	.152	.257	.230	.167	.346
B.I.	.302	−.309	.191	.351	.379	.197	.390	.411	.474	.405	.112	.371	.068
M.C.	.315	−.052	.222	.140	−.185	−.296	.132	.075	−.038	.052	−.041	−.039	.232
D.F.	.414	−.122	.417	.043	.273	.345	.072	.313	.206	.084	−.174	.209	.082
Mash.	.053	.247	.178	−.083	.028	−.178	.027	−.034	−.199	.057	−.376	.032	−.152
Two-Hand	−.016	.276	.142	−.278	−.222	−.273	.135	−.188	−.290	−.291	−.185	−.237	−.050

another as the time of flight instruction increases. Unreliability of the rater might be offered as one explanation, although chance might also be responsible for what appears to be a progressively inverse relationship. It is possible that variation in weather conditions contributed to this relationship.

Check pilots and instructors have been unanimous in stating that one should not expect much agreement between ratings on one check flight and those obtained from any other because of the vast variation in the air conditions under which the student must fly. What might serve as a satisfactory performance on one day might be totally inadequate when weather conditions are different.

Table IV, shows the correlations between the predictors and each of the criteria. A cursory examination of this table indicates that once again none of the correlations were extremely high. Some criteria showed no significant correlations with any test in the battery so at this point an arbitrary decision was made. This decision was to omit from further consideration any criterion which did not have a correlation of at least .340 with any test in the predictor battery. This reduced the number of criteria finally selected for study to thirteen.

In Table V, are the correlations obtained between the individual tests in the predictor battery and each of the criteria selected for further study.

None of the correlations between the tests and the criteria were high. The largest number of comparatively high correlations were obtained between tests in the predictor battery and the C.A.A. over-all grade. The Biographical Inventory showed the largest number of comparatively high correlations with the various criteria, and appeared to be the

most useful predictor as far as data from the check flights were concerned.

The intelligence tests were not highly correlated with the criteria, and the Judgment-Reaction Test generally showed low positive correlations. Both the Test of Aviation Information and the Test of Mechanical Comprehension were correlated most highly with criterion data obtained from the private pilot test, although the correlations with other criteria were low. The Desire to Fly Inventory was likewise correlated highly with the C.A.A. Inspector's data, but its highest correlation was with the rating of Emotional Control. Of the two coordination tests included in the battery, the Two-Hand Coordination Test appeared to have the most significant relationship to the criteria.

Generally, the correlations showed considerable variation and little predictive significance can be attached to them. Although the criteria used differed from those in the present research, much the same type of results were found in the Boston-Midwest Project when correlations were computed between individual predictors and criteria (20).

Table VI shows the partial regression coefficients which give the relative weighting of the nine factors measured by the test battery in predicting the various criteria. The line second from the bottom of the table contains the multiple correlations, which were obtained between the test battery and each criterion. The bottom row contains the corrected multiple correlation coefficients.

In general, no one test was weighted consistently more heavily than any of the other tests in predicting the criteria. As far as criteria obtained from C.A.A. Inspectors were concerned, three tests had

TABLE VI
Beta Weights and Multiple Correlations

		Overall	Land	Emot. Cont.	OSFI Land Scores		Instr. Land Grades		OSFI Land Grades		Slip Dem.	Slip Grades		
					1	3	1	3	2	3	3	1	3	4
Lane	B_1	.067	.226	.292	−.169	.140	.104	−.330	.117	.253	.024	.211	.119	−.122
Otis	B_2	−.306	.072	.390	.383	.325	−.224	.239	−.197	−.434	.094	.356	.027	.156
OSPE	B_3	.003	.109	−.087	−.553	−.756	−.010	−.075	.176	.029	.096	−.136	.062	−.250
A.I.	B_4	.509	−.673	.250	.180	.376	.543	.065	.051	.148	.185	.059	.127	.548
B.I.	B_5	−.049	−.003	−.014	.249	.273	−.022	.401	.298	.373	.269	.135	.272	−.207
M.C.	B_6	.294	.033	.214	.203	−.306	−.397	−.246	.157	.117	.074	.108	−.024	.194
D.F.	B_7	.436	−.325	.292	.099	.277	.449	.112	.223	.049	.093	−.104	.176	.287
Mash.	B_8	.067	.103	.160	.116	.311	.025	.082	.040	−.020	.290	−.521	.205	−.218
Two-Hand	B_9	−.302	.271	−.069	−.386	−.219	−.194	.318	−.358	−.309	−.637	−.056	−.462	−.116
R =		.707**	.655*	.634*	.645*	.622	.688*	.512	.561	.690*	.629*	.579	.546	.550
$_cR$ =		.58	.50	.46	.48	.44	.55	.16	.31	.55	.45	.35	.27	.28

Note: Double asterisk indicates that the correlation is significant at the one per cent level. Single asterisk indicates significance at the five per cent level.

the heaviest weighting. These were The Test of Aviation Information, the Desire to Fly Inventory, and the Two-Hand Coordination Test. Although all the tests were heavily weighted for some of the criteria, the weighting was not consistent. With the possible exception of the Mashburn and the Ohio State Psychological Examination, the test battery showed promise of value in the prediction of the criteria when the weights were properly adjusted for each test in the battery.

The Judgment-Reaction Test proved to be more important in predicting criteria based on performance of specific landing maneuvers than on the over-all grade given by the C.A.A. Inspector. The implication is that whatever factors are measured by the test are involved in landing a plane.

Examination of the multiple correlations at the bottom of Table VI shows that the test battery predicted performance on the C.A.A. private pilot examination far better than it predicted performance on any of the other criteria. This was the only correlation which was significant at the one per cent level. However, seven of the correlations proved to be significant at the five per cent level.

DISCUSSION OF RESULTS

One of the purposes of this research was to examine the possibility of analyzing the factors involved in learning to fly. It appears from an examination of these results that, at least for the group of students studied, and using the criteria which were available in this study, it is possible to discover what some of these factors are. To discover all of the factors involved would be a task far beyond the scope of the present investigation in which the desire was only to demonstrate that such a procedure was feasible.

In general, the corrected correlation coefficients were not sufficiently high to warrant specific conclusions regarding the predictive value of this battery if used with another sample.

It was thought that perhaps the test battery would predict more successfully at different stages in the student's flight training. In other words, it might predict early or ultimate success but, perhaps, not both. Its most successful prediction was for the over-all rating at the end of the course. Other than this, there is no basis for assuming that the battery was of more value in predicting performance at one stage of flight training more than at any other. In addition, there was no consistency in the weighting of the various tests at different time intervals throughout the flight course. Furthermore, it must be remembered that the private pilot test was the only measure which included ratings on more than one type of performance.

There was no consistent relationship between the accuracy of prediction and the number of scores involved in the rating. For example, ratings on Emotional Control were made only once, and a multiple correlation of .634 (significant at the 1% level) was obtained between the test battery and this criterion. Ratings by instructors on landing grades were based on averages of a number of different landings, yet on the sample of landings immediately preceding the third check flight, the multiple correlation was only .512, a correlation that might have arisen by chance.

The fact that there was no consistency in the weighting of items for the various criteria offers several possibilities. First

of all, it may be that each maneuver, even though it is a landing maneuver, actually is a different performance. This might explain differences in weighting between regular landings and landings with a slip. However, on regular landings, there were different weightings for the various tests. Furthermore, landings in different check flights were weighted differently in spite of the fact that instructors were supposedly rating on a fixed scale.

Differences in weights were more apparent when different raters were concerned than when the same rater was involved. This finding is in accord with what has been reported previously in this investigation, namely, that there was little agreement between different raters' ratings of similar performances.

Further evidence that instructors found it difficult to adhere to a fixed scale in grading performance is given by the differences in weighting of items in the instructor's landing grades taken before the first and third check flight. This discrepancy is not so apparent in check pilot's ratings. Such observations are important for the check pilot is in a position to be more objective in his grading than is the instructor who must ride with the student every day.

It is quite possible, however, that much of the inconsistency between ratings on what appear to be similar maneuvers, may be due to weather variations. For example, in quiet air a student might receive a satisfactory grade on landing performance, yet if he were to repeat the same performance in rough air, his grade would be lower. The landing performance in rough air would involve quite different skills than the landing made in still air. This suggests that some method should be devised whereby variations in weather conditions could be included in the rating sheet.

It is impractical to state that there is but one criterion of ability to pilot light aircraft. For a number of years, research on pilot selection was based upon a pass-fail criterion, but this has proved to be of less and less value (25). Rather than to predict success or failure on such a gross basis, a successful battery should have use in predicting performance on specific aspects of flight performance, and it was one of the purposes of this research to demonstrate that certain aspects of flight training might be predicted by a test battery. It appears that the abilities required for the performance of different maneuvers are themselves different. Therefore, one of the uses for this test battery would appear to be in demonstrating students' weaknesses to the instructor before any lessons were given. Furthermore, should a flight course be designed in such a manner that certain specific aspects of flight performance were to be stressed (rather than over-all ability to fly), a battery such as this would be useful in determining applicants' fitness for the course.

SUMMARY AND CONCLUSION

In this research, a battery of nine psychological tests was administered to thirty-seven male flight students. Six of these tests were paper and pencil tests and three were psychomotor tests. Ratings on various aspects of flight performance during a course in primary flight training were available for the sample studied. Intercorrelations among all of the variables were computed, and regression weights and multiple correlations were computed for thirteen selected criteria.

On the basis of the results obtained in this study, it may be stated that:

1. It was possible to assemble a battery of tests which would be of use in analyzing the factors involved in learning to fly light aircraft. Those factors, as measured by tests, were found to be:

Intelligence
Aviation Information
Biographical Information
Mechanical Comprehension
Desire to Fly
Two-hand Coordination
Ability to make Serial Reactions (Eye-hand-foot)
Ability to make rapid judgment reactions.

2. The amount of information about aviation which the flight student has at the beginning of the course seems to be highly related to the grade he receives on the private pilot flight test. For different criteria, however, the various factors measured by the test battery must be weighted differently.

3. Although this test battery produced multiple correlations ranging from .512 to .707 with thirteen separate criteria, the results indicate that the test battery was most successful in predicting when the criterion used was an over-all rating of flight performance, rather than a rating of performance on a specific maneuver.

4. Some writers (8) have expressed doubt regarding the effectiveness of psychomotor tests in the prediction of flight success. Within the limits of this study, it is possible to say that they can be used for this purpose in combination with paper and pencil tests.

5. A new test was included in the battery used in this study. This test was constructed on the hypothesis that it was possible to construct a test which would measure factors related to a specific flight maneuver, in this case, landing. In all but one case, this test contributed to the prediction of performance on this maneuver. It was much more related to the performance of specific maneuvers than to the performance of a series of flight maneuvers combined into a flight test. It is possible that revised methods of scoring this test might produce more fruitful results.

6. In all cases, the use of a battery of tests was of considerably more predictive value than any one of the tests used alone.

7. It was found that trained raters differ from one another in the factors which they stress in evaluating performance, and in any future study attempting to evaluate performance in flying, care should be taken to train raters even more thoroughly in the techniques of rating. Furthermore, some indication of the weather conditions under which each maneuver was performed would be of value in evaluating a student's performance on the same maneuver under different weather conditions.

8. Since ratings made by individuals not having constant contact with flight students seemed to be somewhat more consistent than ratings made by flight instructors who were in constant contact with the student, it seems desirable that relations between the rater and his subject should be kept on as objective a basis as possible.

9. It might be well in future research to use over-all ratings obtained from several sources. In this study, only one over-all measure was available. The use of measures obtained from various sources would make it possible to investigate the possibility that this same battery would predict as successfully, or more successfully, for other over-all ratings. Further-

more, it would seem desirable in future work to use more varied types of specific criteria rather than to study only landing performance. The criteria should be as independent and objective as possible.

10. A further suggestion for future research would be to submit all of the tests included in the battery to such a test selection technique as the Wherry-Doolittle, in order to ascertain if certain of the tests which were found to be unimportant in some of the criteria, might be eliminated from the battery.

11. It is suggested that a battery such as this might be useful in indicating students' potential weaknesses to the instructor so that these students might be eliminated before training, or, if possible, instruction might be directed toward overcoming these weaknesses.

II. The Ability to Perceive and React Differentially to Configurational Changes as Related to the Piloting of Light Aircraft

RONALD R. GREENE
Ohio Wesleyan University

This study was undertaken in an effort to improve the prediction of success in learning to pilot light aircraft. It featured a new psychomotor test which, it was hoped, would contribute materially to the solution of this timely problem.

STATEMENT OF PROBLEM

Procedures in this investigation were planned to test the hypothesis that the ability to perceive and react differentially to visual cues, randomly sampling the visual field, is related to the piloting of light aircraft. This hypothesis was based on several considerations: 1. the failure of simple reaction time as a predictor of success in learning to fly; 2. the comparative success of complicated reaction time as a predictor of flight performance; 3. the observations of the writer, and of other fliers that, (a) correct responding within reasonable time limits seemed more important in piloting than mere quickness of reaction, and that (b) many of the responses required for successful piloting had to be made to indirect visual cues perceived as an inseparable part of the total configuration.

THE INDIRECT VISION TEST

The picture on the following page will be helpful in understanding the Indirect Vision Test Situation. The subject was seated on the adjustable stool in front of the beaver-board screen. Behind the rectangular opening in the center was a tachistoscope that presented a series of simple discrimination problems. This constituted the foveal unit. Behind the small round openings arranged in concentric circles around the center were hooded incandescent bulbs which could be activated in random order. These constituted the parafoveal unit of the apparatus.

The "Instructions to the Subject" further reveal the general nature of the indirect vision test:

This is a test of your ability to see and react to flashes of light at different points in the visual field. When a light flashes on the screen in front of you, you are to place the stick in the corresponding notch in the small board directly in front of you. (Demonstrate)

Use the right hand to move the stick. Hold it in any manner that is convenient for you, but do not grasp it too tightly. Return the stick to the center position immediately after each response so that you will be ready for the next one.

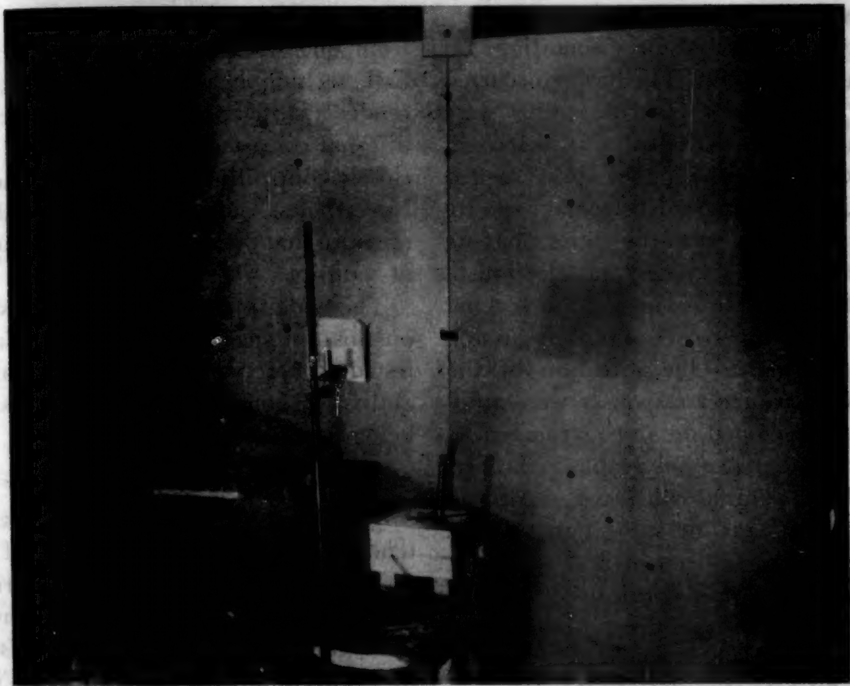
To help you keep looking straight ahead throughout the test, you will look through the small lighted opening in the center of the screen. Behind this opening will appear successive rows of three squares. When the X appears in the middle square, you are to push the button on the table to your left. Use the left hand. Keep the left arm resting on the table and a finger on the button so as to be ready when the X appears in the middle square.

You can ask about anything that you don't understand.

You will be given a practice run of about one minute, after which you will be given the opportunity to ask about anything that it not clear to you.

The test run will continue for about ten minutes.

The sole intended function of the foveal task was that of insuring uniform



INDIRECT VISION TEST FRONT VIEW

fixation during the test and between individuals tested. The problem was reduced to a level of facility that gave, with very rare exceptions, perfect performances. No record was made of responses to the foveal stimuli, but a buzzer signaled that a response was being made.

The foveal unit of the apparatus consisted of a weight-driven, pendulum-controlled, rotating-drum tachistoscope. The tachistoscope was placed behind the screen and the successive stimuli were observable to the subject through a small rectangular opening (1″ x 2″) in the center of the screen. The drum rotated and gave a new exposure every 2 seconds. Illumination of the foveal stimulus was provided by a 20 watt incandescent bulb, hooded, and mounted two inches above and three inches in front of the rotating drum.

The constancy of eye-position was controlled by an adjustable stool and a horizontally fixed headrest 26 inches from the screen.

The stimuli for the indirect vision discrimination task were flashes of light through twenty-one openings in the screen. These openings, three quarters of an inch in diameter, were distributed in circles around the center of the screen at such distances as to give visual angles of 47°, 53°, and 48°. Because of the interference of the response selector the bottom column of lights was deleted, leaving twenty-one openings.

The light flashes were produced by successively activating one or another of the 15 watt bulbs mounted and hooded behind the openings. An alternating current of 10 volts provided the energy for the flashes. The rate and duration of the

flashes were held constant by a high-fidelity electric motor, operating on alternating current. The flashes occurred every 2.2 seconds and the bulbs were activated for a period of .9 of a second. The order in which the lights flashed was controlled by an electromagnetic rotating circuit selector, the pattern having been determined by random selection.

A correct response consisted in placing the stick, within 1.1 seconds, in the notch corresponding to the direction of the light flash. Correct responses were automatically recorded on an electromagnetically operated pen-polygraph. A buzzer, placed in parallel with the recording magnet, signaled the subject that a correct response had been made.

The test ran for a continuous series of 12 cycles of the stimulus pattern. Each cycle flashed 25 lights which made a total of 300 discrete stimulations. Cycles were recorded by an electric counter and were automatically marked on the polygraph tape.

The measure taken as the score was the total number of errors or lights missed.

The screen was of light brown beaverboard, and the general lighting for the test room was provided by an exposed 25 watt light bulb, suspended 3½ feet to the rear and 5 feet above the eye-position of the subject. No daylight was used.

The reliability of the indirect vision test was checked by the split-half and by the test-retest methods. Scores on odd numbered cycles correlated with scores on even numbered cycles .925, (N = 265). The test-retest coefficient of reliability was .784, based on the scores of 130 subjects who took the test a second time after a one-week interval. A likely interpretation of the discrepancy between the coefficients is that the difference is largely due to personal factors which vary from one week to the next. The split-half technique would be expected to render a close approximation of the true internal consistency of the test, by more closely controlling the subjective factors. Nevertheless, it is the lower reliability measure that operates in the determining of validity.

The wide spread distribution of scores on this test suggests that the individuals tested differed markedly in whatever ability or combination of abilities was being used.

OTHER TESTS USED

The design of the experiment was extended to permit observation of the operation of the Indirect Vision Variable when combined with a selected group of pre-flight tests. The tests selected for study in connection with the Indirect Vision Test were:

1. The Self-Administering Test of Mental Ability (Gamma A. M. Otis Quick-scoring)
2. Test of Mechanical Comprehension (Form B, C.A.A.)
3. Desire to Fly (Form XPA)
4. Test of Aviation Information (Form P)
5. Two-Hand Coordination Test

These tests are a sampling of the better paper-pencil and psychomotor predictors now available. The study by Lane shows that all of the old tests in this battery have some relationship to various measures of pilot success. (13)

THE CRITERIA

Our chief interest in this study was the student's ability to fly at or near the end of the normal training period. Consequently most of the criteria selected for use were measures and ratings made during this later period of training.

The only specific maneuver separately

TABLE I
Intercorrelations of Predictors (N = 88)

	Indirect Vision Test	Mental Ability Test	Mechanical Comprehension	Desire to Fly Inventory	Aviation Information Test	Two-Hand Coordination Test
Indirect Vision Test		.039	.103	.264	.025	−.058
Mental Ability Test			.393	.129	.461	.450
Mechanical Comprehension				.169	.395	.498
Desire To Fly Inventory					.068	.083
Aviation Information Test						.199
Two-Hand Coordination Test						

used for validation was that of landing the plane. Landings were selected for two reasons: (1) landing is universally considered one of the crucial aspects of learning to fly, and (2) landing would seem to require the ability to make discriminatory responses to indirect visual cues perceived in relation to the total visual pattern.

The criteria included:

1. C.A.A. Flight Inspectors' Overall Grades
2. C.A.A. Flight Inspectors' Demerit Scores on Landings
3. Check Pilots' Overall Grades on Check Flights 1, 2, 3, and 4. (O.S.F.I.)
4. An Average of the Check Pilots' grades on Landings, 4th Check Flight.
5. Instructors' Ratings of Skill, Judgment, and Emotional Stability (Using the Purdue "Scale for Rating Pilot Competency").

A detailed discussion of these criteria is included in the study by Lane reported in the first part of this monograph.

RESULTS AND DISCUSSION OF RESULTS

The results of this experiment are summarized in the four tables. The correlation coefficients were computed by the product-moment method from raw scores. Table I presents the intercorrelations of the predictors. Inter-correlations among the criteria are presented in Table II. Table III gives the validity coefficients for the predictors when used

TABLE II
Intercorrelation of Criteria (N = 88)

	C.A.A. Overall Grade	C.A.A. Landing Score	O.S.F.I. Overall Check #1	O.S.F.I. Overall Check #2	O.S.F.I. Overall Check #3	O.S.F.I. Overall Check #4	O.S.F.I. Landings Check #4	Purdue Scale Skill	Purdue Scale Judgment	Purdue Scale Emotion
C.A.A. Overall Grade		.583	.227	.119	.229	.507	.298	.446	.165	.270
C.A.A. Landing Score			.170	.128	.220	.291	.162	.206	.145	.145
O.S.F.I. Overall Check #1				.137	.422	.210	.133	.470	.129	.431
O.S.F.I. Overall Check #2					−.075	.382	.420	.385	.088	.190
O.S.F.I. Overall Check #3						.203	−.084	.366	.064	.385
O.S.F.I. Overall Check #4							.756	.542	.241	.474
O.S.F.I. Landings Check #4								.428	.243	.308
Purdue Scale Skill									.407	.692
Purdue Scale Judgment										.354
Purdue Scale Emotion										

TABLE III

Correlations Between Individual Predictors and Criteria (N = 88)

	C.A.A. Overall Grade	C.A.A. Landing Score	O.S.F.I. Overall Check #1	O.S.F.I. Overall Check #2	O.S.F.I. Overall Check #3	O.S.F.I. Overall Check #4	O.S.F.I. Landing Check #4	Purdue Scale Skill	Purdue Scale Judgment	Purdue Scale Emotion
Indirect Vision Test	−.089	.060	.068	.136	−.023	.072	.248*	.261*	.173	.176
Mental Ability Test	−.157	−.109	−.050	.249*	.046	.038	.109	.227*	.312**	.188
Mechanical Comprehension	.279**	−.065	.280**	.287**	.099	.300**	.361**	.376**	.203	.340**
Desire to Fly Inventory	−.012	−.115	.055	.083	.030	−.059	.010	.215*	.186	.153
Aviation Information Test	.195	.054	.206	.338**	.079	.301**	.302**	.283**	.027	.260*
Two-Hand Coordination Test	.006	−.129	.061	.138	−.009	.160	.210*	.206	.271**	.298**

* Significant at the 5% level.
** Significant at the 1% level.

independently. Table IV shows the beta weights for the various tests in relation to each of the ten criteria. The last row of figures in Table IV shows the multiple correlations between the test battery and each criterion.

The Indirect Vision Test shows a marked degree of uniqueness. In only one instance is the inter-correlation above .103. The .103 is with Mechanical Comprehension. The higher one, .264, is with the Desire-to-Fly Inventory. Perhaps those who had the greatest desire to fly were the most highly motivated on the Indirect Vision Test.

Only one other test, the Desire-to-Fly Inventory, shows any considerable degree of independence. The highest inter-correlation for this test is the one with the Indirect Vision Test indicated above.

The four remaining tests of the battery, namely the Mental Ability Test, Mechanical Comprehension, Aviation Information, and the Two-Hand Coordination Test, show a considerable commonality. The inter-correlations for these tests range from .393 to .498. A factorial analysis is needed to appraise the common elements in the hope of reducing the number of tests or test items needed.

A brief visual survey of Table II reveals that the inter-correlations tend to run a little higher for the criteria than they were in the case of the predictors. This was to be expected on several counts. In the first place four of the criteria, the O.S.F.I. check flights, were

TABLE IV

Beta Weights and Multiple Correlations (N = 88)

	C.A.A. Overall Grade	C.A.A. Landing Score	O.S.F.I. Overall Check #1	O.S.F.I. Overall Check #2	O.S.F.I. Overall Check #3	O.S.F..I Overall Check #4	O.S.F.I. Landings Check #4	Purdue Scale Skill	Purdue Scale Judgment	Purdue Scale Emotion
Indirect Vision Test	−.122	.091	.035	.109	−.049	.087	.258	.197	.143	.153
Mental Ability Test	−.380	−.117	−.264	.080	.017	−.224	−.161	.015	.280	−.069
Mechanical Comprehension	.382	−.025	.298	.159	.120	.230	.251	.246	.049	.163
Desire to Fly Inventory	−.006	−.123	.025	.002	.025	−.121	−.108	.103	.102	.064
Aviation Information Test	.236	.139	.209	.239	.049	.299	.252	.157	−.163	.176
Two-Hand Coordination Test	−.007	−.076	−.011	−.019	−.091	.102	.131	.048	.153	.216
R	.472**	.227	.378*	.387*	.138	.425**	.489**	.471**	.422**	.436**

* Significant at the 5% level.
** Significant at the 1% level.

repetitive, and involved two successive ratings by each of the check pilots. Moreover, no serious attempt was made to eliminate overlapping in the selection of criteria. The tenuous nature of the criteria did not warrant selection for mutual exclusion. In one case the rating of a single performance is treated both as a separate criterion and as a factor included in another. This undoubtedly goes a long way toward explaining the comparatively high correlation (.758) between landings of the fourth check flight and the over-all score for the entire flight. The fourth check flight was the only one for which a separate landing grade was used.

It is interesting and perhaps significant that the correlations between check flights involving two ratings by the same check pilot for each of the subjects averages .402, while those between check flights in which each student received one rating from each check pilot averages .170. It is unlikely that differences in performances would account for more than a fraction of this discrepancy. The lower average (.170) suggests unreliability in these ratings, and leads one to wonder whether the higher self-consistency (.402) might not be due to constant errors on the part of the raters. Thus, not only the reliability of the ratings, but the validity as well, is open to question.

The second highest intercorrelation (.692) is found between the skill and emotion aspects of the Purdue Scale. Judgment and emotion correlate .354, while skill and judgment show a relationship of .407. The halo effect may be operating here, since all three aspects are rated by the same instructor. A further investigation of the nature of these relationships would be in order.

Apparently one of the greatest needs in aviation psychology is for more adequate criteria of pilot skill. The photographic method may contribute to the solution, but even this seemingly objective method is not without its problems. Evidence so far suggests that agreement between ratings and photographic records is about the same as between raters. As yet there is no clear way of establishing either of these methods as the one by which the other can be evaluated. A factorial analysis of the data now available is a very necessary next step.

It is obvious from the data in Table III that no single test of the battery is adequate for the prediction of success in learning to pilot light aircraft. Those correlations which differ significantly from zero are marked with one or two asterisks. For those marked with two asterisks the chances are less than one in a hundred that such a correlation would be drawn from a universe where the true value is zero. One asterisk indicates that the odds are between 1 in 20 and 1 in 100 that the true value of the correlation would be zero.

On the whole the correlations are low. They are, on the average, a little lower than those reported by Lane for the same tests. (13) One thing that may help account for this lowering of relationship is the fact that the scores for this study were taken from three quarters of training, rather than the two used by Lane. Changing weather conditions and chance factors have had a greater opportunity to influence the outcomes.

While the Indirect Vision Test did not produce according to expectation, it was shown to have a fairly consistent relationship with the later check pilot grades and the instructor ratings. The highest relationship (.261) is with skill, as estimated by the student's regular instructor,

and the second highest (.248) is with skill in landing, as graded by the check pilot on the fourth check flight. Both of these correlations are significant at the 5% level

The Mechanical Comprehension Test leads the field in predictive value, with seven out of the ten validity coefficients significant at the 1% level, all being above .27. The second-best individual predictor is the Aviation Information Test. It has correlations significant at the 1% level with O.S.F.I. check flights two and four, skill on the Purdue Scale, and with the landing score on the fourth check flight.

The Two-Hand Coordination Test is of little predictive value, except in relation to the landings on the fourth check flight and the three ratings on the Purdue Scale. The Mental Ability Test and the Desire-to-Fly Inventory, according to this study, are of little or no value in predicting pilot success.

It is interesting to note that the Mechanical Comprehension Test is the only one that bears a significant relationship with the C.A.A. examination criteria. Perhaps the validity of these examinations should be seriously investigated, since the granting or withholding of the private pilot license depends upon them.

The beta weights for the Indirect Vision Test reveal that although the original validity coefficients were low they are used almost in toto in the final predictive value of the test battery. The greatest contribution of the Indirect Vision Test is in the prediction of landing ability as measured by the O.S.F.I. technique in the fourth check flight. It may well be that the future usefulness of this test will be highest in relation to this particular maneuver, involving as it does a rapidly changing visual pattern.

The Mechanical Comprehension Test will be seen to make the highest average contribution. Half of the beta weights for this test are above .20. It is especially strong in relation to the C.A.A. overall grade (.382).

The three tests accounting for the greater part of the predictive value of the battery are the Mental Ability Test, the Test of Mechanical Comprehension, and the Aviation Information Test. The predictive value of a battery comprised of these three tests alone would approach that of the six-test battery for most of the criteria. In any case it is recommended that in any future use of these data, tests showing beta weights of less than .10 be dropped from the computation of R for any particular criterion.

A survey of the multiple correlations as presented in Table IV shows that the test battery bears no significant relationship to C.A.A. landing scores or to the over-all scores on the third check flight. For performance during the first and second check flights the battery of tests would seem to have only slight predictive value. The R's for these check flights are .378 and .387. The multiple correlations for the remaining criteria range from .422 (Judgment on the Purdue Scale) to .489 (landings on the fourth check flight), all being significant at the 1% level.

RECOMMENDATIONS

The Indirect Vision test can be improved in several ways. In this experiment the visual cues were presented at a constant rate. Breaking up this rhythm might make the test more differentiating and possibly increase its validity. Again, by modifying the intrument to permit shifting of the positions of the signal lights, a more adequate sampling of the

total visual field could be obtained.

The test might be greatly improved as a predictor of pilot success by using, instead of discrete signals, a continuously changing pattern of stimuli more analogous to that encountered in actual flight. A moving picture of what the pilot sees in flying and landing could be projected upon a translucent screen from the rear. Cues for discriminatory responding would need to be selected and the subject instructed accordingly.

It is recommended that the improved test be correlated with some criterion of success in formation flying. With the rapid growth in aviation, formation flying of some kind in and around airports is becoming a civilian, as well as a military necessity.

The instrument devised for this study has possibilities as a training device, especially if the moving picture technique were to be added. The stimulus control and response selector parts of the apparatus lend themselves to progressive complication, so that a series of tasks at different levels of difficulty could be presented.

A time saving might be effected by determining the smallest necessary sampling on the test. Separate correlations with criteria for each of the cycles, might lead to improved validity as well as a saving in time. On the other hand, perhaps the test should be extended to discover what effect this might have on reliability and validity.

The apparatus can be readily modified to measure simple reaction time, either on the polygraph or with an electric chronometer. By eliminating the indirect stimuli and using only the central unit, simple problems in learning and remembering can be investigated. Without modification the apparatus could be used for studies of learning in relation to distractions. The insertion of a rheostatic control in the stimulus circuit would permit the study of intensity thresholds in the different parts of the visual field. Colored stimuli could also be provided with a minimum of difficulty.

It might be profitable to discover the value or lack of value of the improved Indirect Vision Test in predicting success in automobile driving and in selecting men for industrial positions that seem to involve continuous attending to wide visual areas.

The prime need in the whole program of aviation psychology is for improved measures of success in piloting. It is much easier to point out the need for improved criteria, however, than it is to propose ways and means of bringing about this desired improvement. Already the problem has received a great deal of expert attention. It would seem that a sufficient mass of data is available on a great variety of criteria as to warrant a factorial analysis as a logical next step. Such an analysis would enable us to spot the commonalities among present criteria and to concentrate our further efforts in those areas contributing most to flying success as it has been measured. The results of this analysis would also be helpful as a guide to test making and test improvement.

The introduction of polygraphic and photographic techniques would seem to be in the direction of improved accuracy and objectivity. The little evidence that we have regarding these techniques is not too promising. They have the advantage of rendering a permanent record of what took place, records that can be reviewed again and again. But the interpretation of meaning of the records is still problematical. A control record made by an

expert pilot in the same plane, flying the same course, at nearly the same time, is probably the best basis for comparison. Even this technique does not take into account the capricious nature of minor air currents.

The rating technique is still far from being outmoded. While it is true that ratings generally show an unsatisfactory reliability, there are ways in which the rating of pilot skill can be improved. The training of raters is one method that has not been fully utilized. Training can result in a clearer understanding of the specific behavior being rated, and in a more uniform interpretation of the meaning of the positions along the scale. Multiple ratings on the same flight performance have been previously impossible, inasmuch as practically all of the basic training planes have been two-seaters. With the rapid developments in light plane designing, it would be well to consider the possibility of having two or more raters fly simultaneously with the subject and make individual evaluations of the various maneuvers.

With the large number of criteria sampling the student's skill at successive points in the training period, one is prompted to raise the question as to whether or not some of the early criteria might have predictive value for later performance. The intercorrelations of the criteria in this study are a step in the direction of finding an answer to this question. It is recommended that the results of this experiment, along with the data from Lane's study, be used in determining the value of measures of early flight performance in the prediction of ultimate success or failure in learning to pilot light aircraft.

SUMMARY

The experiment described in detail in the preceding chapters was addressed to the general problem of improving prediction of success in piloting light aircraft.

The specific problem was that of determining the relation between the ability to perceive and react differentially to configurational changes and the ability to pilot light aircraft. A test was devised to measure the ability to perceive the total visual field and to make selective responses to changes therein. The changes were provided by flashes of light sampling the visual field. The intensity and duration of these stimuli were below the threshold for after-images. Correct directional responses were recorded electromagnetically on a pen-polygraph. A simple foveal task was employed to control eye-fixation. Indirect visual cues were presented at the rate of one every 2.2 seconds. To be recorded the correct response had to be made within 1.1 seconds after the stimulus occurred. Reliability coefficients for this test were satisfactory for its use.

Scores on this test were correlated with ten criteria of success in flying. The r's are low, but compare favorably with the predictive value of other pre-flight tests. Inter-correlations between the Indirect Vision Test and five more-or-less widely used pre-flight tests reveal a minimum of over-lapping. The beta weights assigned to the Indirect Vision Test, when all six predictors are used as a battery, further substantiate its unique contribution.

The R's between the battery of tests and the various criteria are comparable to those of similar batteries. They are sufficiently high and significant in rela-

tion to eight out of the ten criteria as to warrant use of the battery in practical situations, pending the development of better predictors.

The correct evaluation of predictors is still problematic, due to weaknesses in the criteria of success in flying. It is possible that the Indirect Vision Test, and the battery of which it is a part will show higher validity when related to more adequate criteria. On the other hand, the validity coefficients may be lowered by the improvement of criteria.

Insofar as the criteria used in this study are true measures of pilot ability, performances on the Indirect Vision Test and on the test battery have something in common with the successful piloting of light aircraft.

Suggestions have been made for improving the Indirect Vision Test and for investigating its usefulness in fields other than aviation. It has also been pointed out that the results of this experiment provide a basis for the investigation of additional problems in aviation psychology.

BIBLIOGRAPHY

1. Carlson, W. A. Intelligence testing of flying cadet applicants; a report on psychometric measurement. *J. Aviat. Med.*, 1941, *12*, 226-229.
2. Delucchi, J. R. Selection of military aircraft pilots; our viewpoints about professional selection and psychological examinations. *J. Aviat. Med.*, 1942, *13*, 234-244.
3. Dockeray, F. C., and Isaacs, S. Psychological research in aviation in Italy, France, England, and the American Expeditionary Forces. *J. comp. Psychol.*, 1921, *1*, No. 2, 115-148.
4. Edgerton, H. A., and Walker, R. Y. History and development of the Ohio State Flight Inventory. Part I, Early versions and Basic research. C.A.A., Div. of Research, Report No. 47, Washington, D.C., July 1945.
5. Festinger, L., and Kogan, Odbert, and Wapner. An analysis of inspector's ratings of check flight as recorded on form CAA-3422. C.A.A., Div. of Research, Report No. 58, Washington, D.C., March, 1946.
6. Jenkins, J. G. Selection and training of aircraft pilots, *J. consult. Psychol.*, 1941, 8, 228-234.
7. Johnson, H. M., and Boots, M. L. Analysis of ratings in the preliminary phase of the C.A.A. training program. C.A.A., Div. of Research, Report No. 21, Washington, D.C., October, 1943.
8. Johnson, H. M., and Boots, M. L. On the actual and potential value of biographical information as a means of predicting success in aeronautical training. C.A.A. Airman Development Div., Report No. 32, Washington, D.C., August, 1944.
9. Kelley, E. L. The development of "A scale for rating pilot competency." C.A.A., Div. of Research, Report No. 18, Washington, D.C., July, 1943.
10. Kelley, E. L., and Johnson, A. P. Analysis of test data on 1938-39 C.A.A. students. In: Studies of predictors of achievements in learning to fly. Four early investigations of aptitude and personality tests in relation to flying. C.A.A., Div. of Research, Report No. 27, Washington, D.C., March, 1944.
11. Kellum, W. R. Recent developments in selection of candidates for aviation training. *Amer. J. Psychiat.*, 1943, *100*, 80-84.
12. Kogan, Wantman, and Dunlap, J. Analysis of the Desire to Fly (D-F) inventory. C.A.A., Div. of Research, Report No. 50, Washington, D.C., October, 1945.
13. Lane, G. G. Prediction of success in learning to fly light aircraft (unpublished).
14. Liljencrantz, E. Problems in the selection of aviators. *J. Aviat. Med.*, 1942, *13*, 107-120.
15. McFarland, R. A. and Channell, R. C. A revised serial reaction time apparatus for use in appraising flying aptitude. C.A.A., Airman Development Div., Report No. 34, Washington, D.C., September, 1944.
16. McFarland, R. A., and Channell, R. C. A revised two-hand coordination test. C.A.A., Airman Development Div., Report No. 36, Washington, D.C., October, 1944.
17. McFarland, R. A., and Franzen, R. The Pensacola study of naval aviators; final summary report. C.A.A., Div. of Research, Report No. 38, Washington, D.C., November, 1944.
18. N.R.C. Committee on Selection and Training

of Aircraft Pilots. An historical introduction to aviation psychology, C.A.A., Div. of Research, Report No. 4, Washington, D.C., October, 1942.

19. N.R.C. Committee on Selection and Training of Aircraft Pilots, history and development of the Ohio State Flight Inventory, Part II. Recent version and current application, C.A.A., Div. of Research, Report No. 51, Washington, D.C., November, 1945.

20. N.R.C. Committee on Selection and Training of Aircraft Pilots. Report on the Boston-Midwest project. C.A.A., Div. of Research, Report No. 52, Washington, D.C., November, 1945.

21. N.R.C. Committee on Selection and Training of Aircraft Pilots. The C.A.A. National Testing Service; Summary of test results and comparisons with success in flight training. C.A.A., Div. of Research, Report No. 39, Washington, D.C., November, 1944.

22. Peters, C. C., and Van Voorhis, W. R. *Statistical procedures and their mathematical bases.* McGraw-Hill Book Co., New York and London, 1940.

23. Staff, Psych. Research Project (Pilot) Psychological Research in pilot training in the A.A.F. *Amer. Psychol.*, 1946, *1*, No. 1, 9-16.

24. Toops, H. A. (Chairman). Annual report of the Committee on Technical Research. Ohio College Association Bulletin, 1941-42, Seventeenth, No. 130, p. 2501.

25. Viteles, M. S. The aircraft pilot, Five years of research; A summary of outcomes. C.A.A., Div. of Research, Washington, D.C., June 15, 1945.

26. Wapner and Bakan. The agreement between ratings of one inspector as recorded on the Ohio State Flight Inventory and instrument readings obtained from photographic records, Progress report, N.R.C. Committee on Selection and Training of Aircraft Pilots, April 2, 1946. (Unpublished)

27. Wapner, S., Odbert, H. S., Kogan, L. S., Festinger, L. The development of a civilian key for the inventory of personal data for prospective pilots. (B-I) NRC Committee on Selection and Training of Aircraft Pilots. October 5, 1945.

28. World Book Co. *Test of mental ability,* C.A.A. Copyright, 1937.

ACKNOWLEDGMENT

ACKNOWLEDGMENT with gratitude is made to Dr. Samuel Renshaw both for direction of the study and for generous access to laboratory accommodations and funds.